Blind Inpainting with Object-aware Discrimination for Artificial Marker Removal

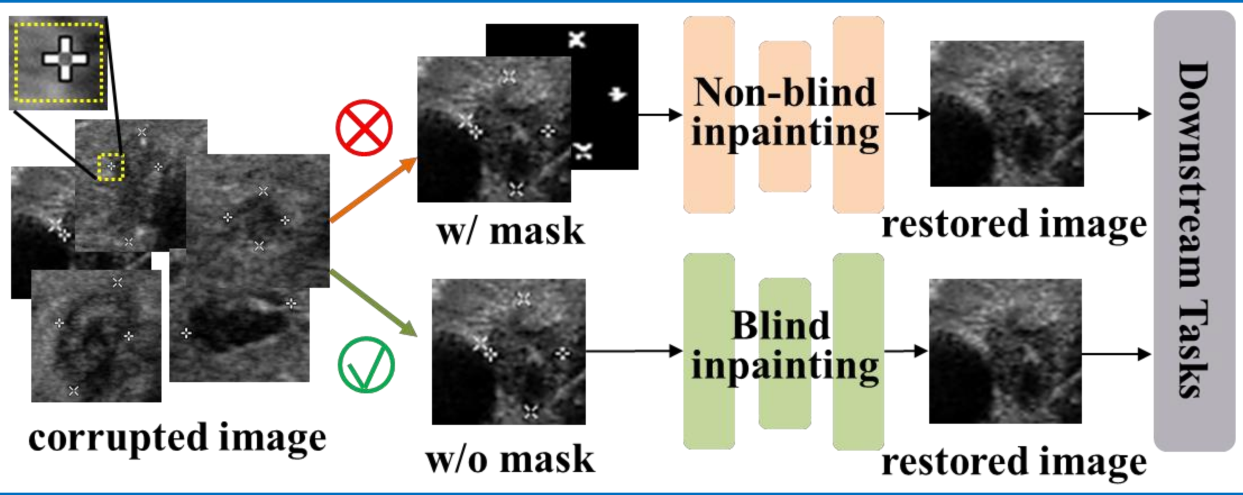
Xuechen Guo[1], Wenhao Hu[1], Chiming Ni[1], 1$Wenhao Chai[2], Shiyan Li[3], Gaoang Wang[1,✉]

1. ZJU-UIUC Institute, Zhejiang University, China
2. Department of Electrical & Computer Engineering, University of Washington, USA
3. Sir Run Run Shaw Hospital, Zhejiang University, China

1. MOTIVATION

Reconstruction for corrupted medical images:

- Medical images often incorporate doctor-added markers that can hinder AI-based diagnosis. This issue highlights the need of inpainting techniques to restore the corrupted visual contents.
- Existing methods require manual mask annotation as input, limiting the application scenarios.
- In this paper, we propose a novel **blind inpainting** method that automatically reconstructs visual contents within the corrupted regions without mask input as guidance.



2. RELATED WORK

Existing works

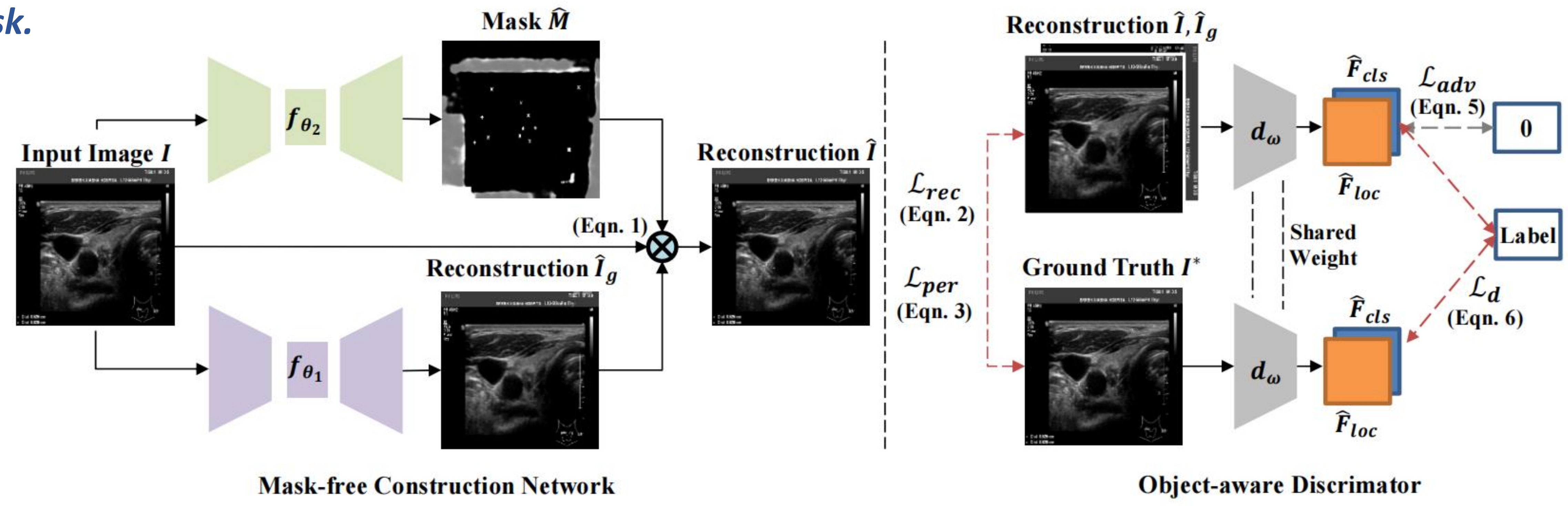
- **Inpainting methods for image completion:** gated convolution-based, transformer-based, diffusion-based methods
- **Extensive applications in medical imaging:** chest X-ray, brain MRI,… → involves **manual mask annotation**, inconvenient, time-consuming, and error-prone.
- **Blind inpainting methods**: a more practical solution, mask-free.→ still **fail** to localize corrupted regions, leading to sub-optimal solutions in reconstruction.

3. PROPOSED APPROACH

An efficient mask-free network for the blind inpainting task.

- **A two-branch reconstruction network :** f_θ guides the inpainting process to focus on corrupted regions, which are unknown to model. Branch $f_{\theta 1}$ is for inpainting missing content in corrupted regions localized by $f_{\theta 2}$. Each branch utilizes the same upsampler - convolution - downsampler structure based on gated convolution, but is with distinct parameters.
  → Eliminate dependency on a manual mask input.
- **An object-aware discriminator :** Utilize and enhance a dense object detector such as YOLOv5 to build our discriminator, thus to accommodate markers of different relative sizes in corrupted images. This leverages the detector's powerful recognition capabilities for pixel-based classification in local regions.
  → Enhance adversarial training.

Thus, our end-to-end blind inpainting model can produce reconstructions closely resembling clean images.



$$\hat{I}_g = f_{\theta_1}(I),$$
$$\hat{M} = f_{\theta_2}(I),$$
$$\hat{I} = \hat{M} \odot \hat{I}_g + (1 - \hat{M}) \odot I,$$
$$\mathcal{L}_{\text{rec}}(\theta) = \|I^* - \hat{I}_g\|_1 + \|I^* - \hat{I}\|_1,$$
$$\mathcal{L}_{\text{per}}(\theta) = \|\phi(I^*) - \phi(\hat{I}_g)\|_2 + \|\phi(I^*) - \phi(\hat{I})\|_2,$$

$$\hat{F}^{\Omega}_{\text{cls}}, \hat{F}^{\Omega}_{\text{loc}} = d_\omega(\Omega), \quad \Omega \in \{I^*, \hat{I}_g, \hat{I}\},$$
$$\mathcal{L}_{\text{adv}}(\theta) = -\mathbb{E}_{\Omega \in \{\hat{I}_g, \hat{I}\}} \log(1 - \hat{F}^{\Omega}_{\text{cls}})$$
$$\mathcal{L}_{\text{d}}(\omega) = \sum_{\Omega \in \{I^*, \hat{I}_g, \hat{I}\}} \mathcal{L}_{\text{cls}}(\hat{F}^{\Omega}_{\text{cls}}; \omega) + \mathcal{L}_{\text{loc}}(\hat{F}^{\Omega}_{\text{loc}}; \omega).$$
$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{rec}}(\theta) + \lambda_2 \mathcal{L}_{\text{per}}(\theta) + \lambda_3 \mathcal{L}_{\text{adv}}(\theta) + \mathcal{L}_d(\omega)$$

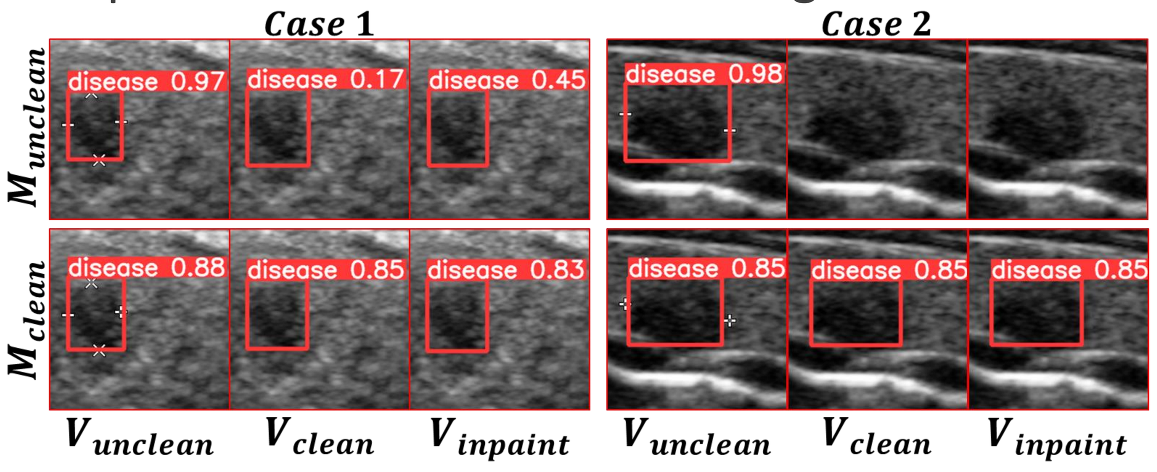
4. EXPERIMENTS

Datasets

- **US:** ultrasound, from Sir Run Run Shaw Hospital
- **MRI**: magnetic resonance imaging, from Prostate MR Image Segmentation Challenge
- **CM**: electron microscopy, from the MICCAI 2015 gland segmentation challenge

Motivation Verification

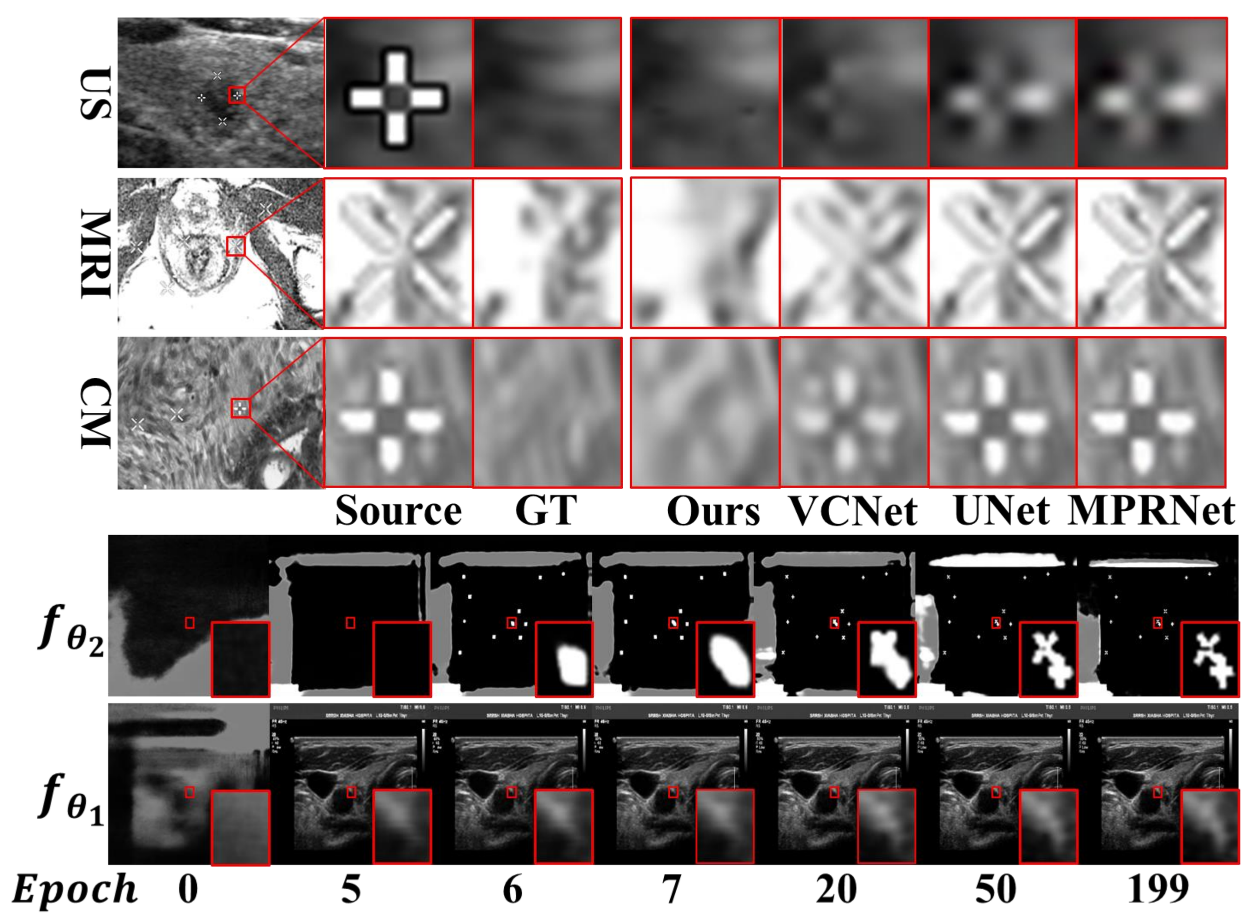
- **Verify the motivation by YOLOv5 for lesion detection on US dataset**: Train YOLOv5 models $M_.$ on unclean data with artificial markers and clean data respectively. $V_.$ is test sets. Inpainting $V_{unclean}$ by our model to obtain $V_{inpaint}$.
- $M_{unclean}$ detects lesions relying on marker recognition, rather than understanding medical semantics as M_{clean}. It proves the negative impact of unclean data on AI diagnostics.



Models	Test sets	P	R	mAP@.5	mAP@.5:.95
$M_{unclean}$	$V_{unclean}$	0.875	0.860	0.860	0.844
	V_{clean}	0.500	0.594	0.556	0.248
	$V_{inpaint}$	0.583	0.429	0.511	0.221
M_{clean}	$V_{unclean}$	0.780	0.754	0.773	0.442
	V_{clean}	**0.770**	**0.696**	**0.734**	**0.425**
	$V_{inpaint}$	**0.664**	**0.719**	**0.676**	**0.389**

Performance & Comparisons with SOTA

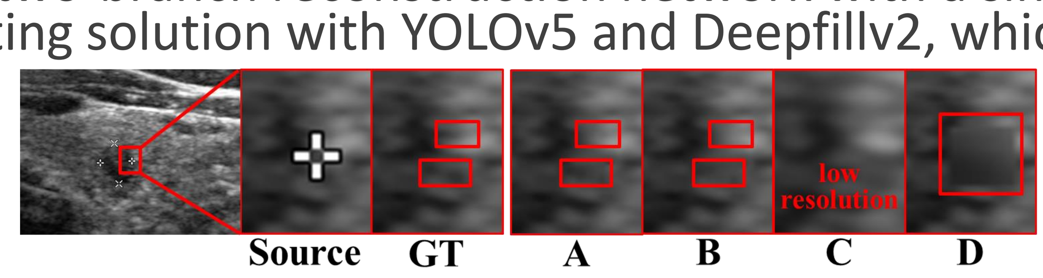
- **Models for comparisons:** MPRNet, Unet, VCNet and our proposed model
  Metrics(mean±s.d). In parentheses are metrics further calculated only within mask areas.
- Ours generates visually appealing results. Other models exhibit varying levels of restoration failure.



Data	Methods	PSNR ↑	SSIM ↑	MSE ↓
US	MPRNet	$37.877_{\pm 3.289}$ (13.478)	$0.995_{\pm 0.002}$ (0.429)	$13.027_{\pm 10.201}$ (3213.933)
	UNet	$35.262_{\pm 1.319}$ (14.899)	$0.985_{\pm 0.004}$ (0.419)	$20.499_{\pm 9.442}$ (2280.374)
	VCNet	$36.891_{\pm 1.425}$ (28.988)	$0.971_{\pm 0.012}$ (0.801)	$14.442_{\pm 6.910}$ (**87.293**)
	Ours	$\mathbf{47.673}_{\pm 5.415}$ (**30.016**)	$\mathbf{0.999}_{\pm 0.001}$ (**0.855**)	$\mathbf{2.633}_{\pm 5.856}$ (103.111)
MRI	MPRNet	$34.860_{\pm 1.992}$ (17.692)	$0.991_{\pm 0.001}$ (0.627)	$23.298_{\pm 9.599}$ (1226.490)
	UNet	$29.736_{\pm 2.004}$ (18.021)	$0.961_{\pm 0.012}$ (0.625)	$75.659_{\pm 29.296}$ (1003.576)
	VCNet	$31.315_{\pm 1.405}$ (21.117)	$0.947_{\pm 0.029}$ (0.705)	$63.405_{\pm 18.734}$ (423.108)
	Ours	$\mathbf{40.049}_{\pm 7.004}$ (**26.159**)	$\mathbf{0.994}_{\pm 0.003}$ (**0.821**)	$\mathbf{7.153}_{\pm 9.627}$ (**203.967**)
CM	MPRNet	$35.184_{\pm 1.368}$ (18.354)	$0.991_{\pm 0.002}$ (0.702)	$20.505_{\pm 6.460}$ (1004.690)
	UNet	$34.239_{\pm 0.847}$ (19.472)	$0.984_{\pm 0.001}$ (0.707)	$24.881_{\pm 4.931}$ (1015.378)
	VCNet	$32.230_{\pm 0.350}$ (22.268)	$0.956_{\pm 0.007}$ (0.718)	$39.016_{\pm 3.098}$ (387.710)
	Ours	$\mathbf{41.419}_{\pm 1.902}$ (**28.437**)	$\mathbf{0.997}_{\pm 0.001}$ (**0.839**)	$\mathbf{2.595}_{\pm 1.284}$ (**165.442**)

Ablation study on US dataset

- **A**: our complete model. **B**: our object-aware discriminator with the one in Deepfillv2.
  C: our two-branch reconstruction network with a single branch one. **D**: a two-stage non-blind inpainting solution with YOLOv5 and Deepfillv2, which are the basis of our implementation.



Type	PSNR ↑	SSIM ↑	MSE ↓
A	$\mathbf{47.673}_{\pm 5.415}$	$\mathbf{0.999}_{\pm 0.001}$	$\mathbf{2.633}_{\pm 5.856}$
B	$33.283_{\pm 2.023}$	$0.984_{\pm 0.006}$	$33.948_{\pm 16.306}$
C	$29.306_{\pm 2.131}$	$0.883_{\pm 0.038}$	$87.551_{\pm 52.855}$
D	$43.551_{\pm 3.014}$	$0.998_{\pm 0.001}$	$4.583_{\pm 9.094}$

5. CONCLUSIONS

Contributions

- We propose a novel blind inpainting method with a mask-free reconstruction network and an object-aware discriminator for artificial marker removal in medical images.
- Eliminate dependency on the manual mask input for corrupted regions in an image.
- Practicability of employing an dense object detector to the discriminator.
- Efficiency and robustness on multiple medical image datasets such as US, EM, and MRI.

Future work

- Combine diffusion models in the reconstruction network
- Validate the performance in large hole blind inpainting